

DATA SCIENCE TO PROMOTE CORPORATE GOVERNANCE OF INFORMATION TECHNOLOGY

Pedro Solana-González, Universidad de Cantabria, pedro.solana@unican.es

Adolfo Alberto Vanti, Universidade Regional Integrada do Alto Uruguai e das Missões (URI), adolfo.vanti@gmail.com

Rosane Maria Seibert, Universidade Regional Integrada do Alto Uruguai e das Missões (URI), rseibert@san.uri.br

RESUMO

A teoria e prática de Data Science pode ser operacionalizada de diferentes maneiras, mas é a técnica de Data Mining uma das mais utilizadas atualmente pelas instituições pois permite descobrir e gerar conhecimento sobre as bases de dados que possuem. Mais que isso, esse enfoque de mineração de dados integrado a conhecimento de especialista pode promover o aprendizado de máquina que neste trabalho ajuda a Governança de TI para melhorar a Prestação de Contas. Para tal se utilizou uma base de dados constituída junto a quase seiscentas empresas com avaliação consolidada de maturidade de quatro dezenas de processos tecnológicos para uma análise didática de uso e aproximação dessa técnica ao contexto de gestão.

Palavras-chave: Prestação de contas; Governança de TI; Ciência de dados; Mineração de dados.

ABSTRACT

Data Science theory and practice can be operationalized in different ways but is the Data Mining technique that is one of the most used by institutions today because it allows to discover and generate knowledge about the databases they have. More than that, this data mining approach integrated with expert knowledge can promote the machine learning that in this work helps IT Governance to improve Accountability. For this purpose, a data base was set up with almost six hundred companies with a consolidated maturity evaluation of four dozen technological processes for a didactic analysis of use and approximation of this technique to the management context.

Keywords: Accountability; IT governance; Data Science; Data Mining.

1. INTRODUÇÃO

A terminologia Data Science (DS) se incorporou rapidamente ao dia-a-dia de organizações e se confunde muitas vezes com o próprio Data Mining (DM). Conforme Provost e Fawcett (2016, pag. 2) DS é um conjunto de princípios fundamentais que norteiam a extração de conhecimento a partir de dados, enquanto que DM é a extração de conhecimento a partir deles por meio de tecnologias que incorporam esses princípios. Assim, o processo de tomada de decisão efetivamente se dá em dados e não em instituição do decisor, porém de maneira que antecipe situação, antecipe o que vai ocorrer, que se possa prever o comportamento de consumo, de finanças e de diversos tipos de situações dentro da empresa (modelos preditivos que auxiliam o processo de decisão em segundos).

Data Science pode promover a Governança de TI (GTI), principalmente quando alcança os princípios de Governança Corporativa (GC) como Transparência, Redução de Riscos, Conformidade e Prestação de Contas. Esse trabalho se direciona à Prestação de Contas ou *Accountability* em que a análise preditiva de dados permite que promova a GTI e consequentemente a GC, convertendo-se assim em um ativo estratégico.

A promoção da Governança de TI pode ser ocorrer de diferentes maneiras no auxílio da GC mas é na maturidade dos processos tecnológicos que ela pode ser analisada em conjunto com outros processos como os relacionados ao planejamento de TI, contratação de terceiros para implantação de sistemas, riscos operacionais, segurança de sistemas e até performance dos funcionários na área tecnológica e de gestão. O *framework* da família Cobit em todas as suas versões atende bem a promoção da Governança de TI quando se posiciona que o estabelecimento estrutura de governança envolve estruturas, processos, funções e responsabilidades que asseguram variados investimentos convergentes com as estratégias e objetivos organizacionais.

Assim este trabalho formulou o seguinte problema de pesquisa: Como promover Governança de TI com uso intensivo da Data Science? Para responder esse problema de pesquisa se utilizou de enfoque de Ciência de Dados para promover a GTI usando Data Mining que, de maneira didática, se busca aproximar mais esse tipo de técnica ao contexto de gestão. O trabalho apresentado está estruturado da seguinte forma: em primeiro lugar, avança na revisão de literatura que trata da Governança de TI no contexto da Governança Corporativa; Em segundo lugar, expõe-se a metodologia que suporta a promoção da Governança de TI, passando a apresentar o processo de mineração de dados e as etapas específicas da investigação, para posteriormente aproximar os resultados do trabalho e finalizar com as conclusões.

2. REVISÃO DA LITERATURA

Para a revisão da literatura se contemplou temas relacionados à Data Science, Mineração de Dados, Governança Corporativa, Prestação de Contas e Avaliação de Processos Tecnológicos com CobiT.

Governança de TI à luz da Governança Corporativa

A Governança de TI (GTI) consiste em direitos de decisão e de matriz de responsabilidades para motivar comportamentos desejáveis no uso da TI, ampliando-se também para funções de supervisão, monitoramento, controle e direção das

organizações (WEILL e ROSS, 2005; ITGI, 2003). Esta governança corporativa de TI pode ser estruturada através do Cobit (ISACA, 2018), o qual atende à todas as partes interessadas da organização.

Para entender a integração desta GTI com a GC, torna-se necessário contextualizar melhor o cenário em que atuam princípios ou requisitos de Transparência, Assimetria da Informação, Conflitos de Agência e Prestação de Contas. Do contrário, se a GTI não trabalhar para uma melhor gestão e governança da empresa ela se tornará limitada e assim dá razão à críticas de tecnologia por tecnologia, o que ocorre muito nas empresas proporcionando dessa maneira problemas que podem inclusive pôr em risco a sobrevivência das mesmas pois a cada dia a informação de qualidade é base de todo o conhecimento e inovação, está se tornando o ativo mais estratégico na informação.

Então, a GC e o aumento do uso de tecnologia global nos negócios exigem um aumento dos níveis de maturidades empresariais e para tal, processos tecnológicos permitem estabelecer uma relação direta entre qualidade da informação e aumento da transparência com redução de assimetria da informação (LEE, LEE e WANG, 2017; AKBAR; et al., 2016; ALHURAIBI, 2017) e possibilidade de redução de conflitos de agência (CAMPREGHER e LONGONI, 2017; OCDE, 2015), bem como foco deste trabalho uma melhor Prestação de Contas. Aumenta a credibilidade, a confiança, a responsabilidade e a previsibilidade da formulação de políticas organizacionais, facilitando o monitoramento dos *stakeholders* (SEIBERT e MACAGNAN, 2017; SHAMBAUGH e SHEN, 2018).

A GC se formaliza de diferentes maneiras e customizações empresariais para atender seguir bem com as legislações específicas, conformidade legal mas também se formaliza pelo código das melhores práticas de GC do Instituto Brasileiro de Governança Corporativa (IBGC, 2009), derivado de OECD (2015) e que se torna bem mais relevante quando da adequada integração das bases de dados (GIRALDO, JIMÉNEZ e TABARES, 2017), mais ainda quando essas grandes bases de dados possam gerar conhecimento, melhores decisões e novos comportamentos de gestão que atendam ainda melhor a sociedade em geral.

A distância entre os interesses dos *stakeholders* e os da organização precisa ser gerenciada de maneira mais autônoma para que a empresa seja mais transparente e preste melhor suas contas com alcance a uma plenitude de melhores práticas em GC (BRANDES e DARAI, 2017; OXELHEIM, 2018), diminuindo assim a assimetria da informação, custos e riscos (SHAMBAUGH e SHEN, 2018) e consequentemente os conflitos de agência afetando positivamente a performance organizacional essa qualidade da informação (STOUTHUYSEN, SLABBINCK e ROODHOOFT, 2017) e a adição de valor à empresa (HIEBL, 2018).

A empresa opera fundamentalmente em sistemas de informações e tecnologias associadas e por isto a insistência que boas práticas de GTI podem aumentar as garantias de GC com seus conjuntos de princípios e práticas. Assim que, neste estudo, se focaliza nessas práticas e se converte as mesmas em um *framework* de avaliação de processos tecnológicos que converte metas organizacionais em metas de TI e logo metas corporativas. Isso se dá com o Cobit (ISO/IEC, 2018); De Haes e VAN GREMBERGEN, 2015) como GTI que totaliza se apresenta com metas corporativas integradas às metas de TI, processos e atividades os quais estabelecem indicadores (ISACA, 2012; CARVALHO, ROMÃO, FAROLEIRO, 2016) que controlam os princípios de GC e também de confidencialidade, integridade e disponibilidade da

informação, requisitos estes coerentes com os princípios do Comitê dos Pronunciamentos Contábeis (CPC).

3. METODOLOGIA NA PROMOÇÃO DA GOVERNANÇA DE TI

A metodologia utilizada no modelo computacional que avaliou a promoção da GTI em processos tecnológicos baseado em Cobit foi Data Mining e operacionalizando-se com a ferramenta computacional Weka de código aberto. Assim a mineração de dados permite identificar padrões, prever situações (WAN et al., 2016) e realizar aprendizado de máquina entre outros recursos de igual robustez, combinando técnicas estatísticas, inteligência artificial e regras de negócios (FARAZZMANESH e HOSSEINI, 2017).

Cada um desses aspectos é avaliado de acordo com seu grau de maturidade em 6 níveis:

0 – Inexistente: Neste nível há uma absoluta falta do processo. A organização não tem conhecimento sobre as implicações que a falta do processo pode gerar.

1 – Inicial: Neste nível os processos são esporádicos e desorganizados, não existe documentação e controle alguma.

2 – Repetitivo, mas intuitivo: Neste nível os processos seguem um padrão de regularidade, com alta dependência do conhecimento dos indivíduos.

3 – Definido: Neste nível os procedimentos estão estabelecidos e são cumpridos. Início do uso de indicadores para controle.

4 – Gerenciado: Neste nível os processos estão integrados e alinhados. As metas e planos são baseados em dados e indicadores consistentes.

5 – Otimizado: Boas práticas são seguidas e automatizadas, com base em resultados de melhoria contínua.

A partir dessas variáveis que medem o grau de maturidade em 34 processos da organização, os níveis de TI foram classificados de acordo com 548 organizações para os 34 processos Cobit. Como isso se construiu um modelo computacional e um enfoque matemático de descobrimento de conhecimento com métodos de *Machine Learning* e *Data Mining* (MITCHELL, 1997; MCCUE, 2007) puderam atuar para entender como se pode promover a Governança de TI em empresas. Ou seja, para exemplificar, se conheço como se conecta em forma de árvore de decisão os processos tecnológicos, posso melhorar meu nível de investimento em TI.

Prestação de contas ou responsabilizações em atos financeiros (NICOL, 2018) se agravaram nos últimos anos principalmente devido às crises econômicas e ao processo de legitimação governamental (KRAFT e WOLF, 2018) em que muitas vezes essa é direcionada nível mais operacional para justificar problemas de falhas de sistemas e de comportamentos éticos (NEI, FOSTER, NESS e NEI, 2018).

Em Ferry e Murphy (2018) o tema de prestação de contas ou responsabilização se aliou à transparência no intuito de construir um sistema público democrático mais robusto, ou mesmo em Thomann, Hupe e Sager (2018) em que usando um enfoque de configuração foi possível aliar atuação institucional com inspeções de específico setor. Assim se proporciona uma cultura importante em saber o que realmente ocorre na empresa baseado em dados e no que os mesmos podem nos dizer aliado ao conhecimento de especialistas, uma cultura de *self-accountability* e *self-regulatory*

(DHIMAN, SEM e BHARDWAJ, 2018) em padrões de liderança estratégica 24x7 que possam cada vez mais serem aliadas a uma boa gestão do conhecimento (NAJMI, KADIR e KADIR, 2018).

Auditorias (ROMZEK e INGRAHAM, 2000) com responsabilizações baseadas em bases de dados que compõem as atividades institucionais são mais efetivas pois Data Mining pelo reconhecimento de padrões de comportamento de dados gerados por algoritmos (FAYYAD, SHAPIRO E PADHRAIC, 1996) e que pode também ser de maneira preditiva (TAN, STEINBACH e KUMAR, 2005), possibilitam incorporar modelos de inteligência artificial a fim de prever situações de maior prestação de contas e de menor risco institucional (BAJO et al., 2012) bem como melhorar efetivamente o processo decisório.

Conforme Shmueli e Koppius (2011) o poder preditivo com precisão se refere à capacidade de um modelo de gerar cenários futuros observações que não foram incluídas na amostra original, porém com merecido cuidado no trato das diferenças entre informação e conhecimento dessa *accountability* corporativa (DU RIETZ, 2018) e proativa (ALOM, 2018) que gera também maior transparência, ética (ODONGO e WANG, 2018) e privacidade (PAYNE, LANDRY e DEAN, 2015).

Prestação de Contas, responsabilização ou *accountability* então é requisito essencial de uma boa governança e a mesma somente é bem evidenciada quando os dados e informações refletirem adequadamente as atividades, gerando significativo conhecimento para as auditorias e para melhores decisões (SURYANI, SASMITA e PURNAWAN, 2015; GANTMAN e FEDOROWICZ, 2016), integrando dessa forma de maneira interdisciplinar evidências contábeis de bases de dados com enfoque de prestação de contas (ATKINS e MAROUN, 2018).

Uma estrutura tecnológica e de análise ou avaliação de dados permite que se tenha um maior controle sobre as atividades da empresa, implementando processos, estruturas e mecanismos relacionados, conduzindo assim à comportamentos desejáveis de seus executivos, e que em consequência, se gere valor nas atividades e com percepção à sociedade. Conduzir a comportamentos desejáveis nesse enfoque é proporcionar avaliações ou auditorias futuras com predição (SHIRI, AMINI e RAFTAR, 2012) de ocorrências através de aquisição de conhecimento e aprendizado de máquina.

Este trabalho implementado com *software* Weka e derivações modelaram regras caracterizadas com Data Mining (WITTEN e FRANK, 2005) em algoritmos em que uma pró-atividade (ALOM, 2018) em melhores adequações de *accountability* podem ser alcançadas numa aplicação de níveis consolidados de maturidade de processos tecnológicos, estes baseados em dados formados a partir de percepções de avaliações de diferentes profissionais integrados a uma avaliação de conhecimento com especialista em classificação de níveis de transparência. Logo essas avaliações são tratadas em conjunto considerando por exemplo técnicas de Árvores de Decisão (C4.5), Naive Bayes (VALLE, VARAS e RUZ, 2012) Multi-layer Perceptron (MLP) e Lazy Learning k-NN (LÓPEZ DE MANTARAS E ARMENGOL, 1998; WU, KUMAR et al. 2008; CHOI et al., 2017).

Esse trabalho também contribui indiretamente para a para a redução comportamentos burocráticos em contextos empresariais (RYU, CHANG, 2017), (FILGUEIRAS, 2018, p. 73), aumentando a racionalidade e melhorando sua governança pois assim pode-se saber com maior exatidão em que processo tecnológico a instituição deverá investir com maior robustez. Uma governança deve

se ajustar a comportamentos de gestores e também proporcionar que estes tenham um comportamento adequado ao que se espera de uma gestão sustentável. Também uma boa governança deve seguir leis para apresentar uma adequada gestão e prestação de contas em que o enfoque de design pode melhorar o processo de geração de *accountability* e até mesmo aprimorar o enfoque orçamentário (CUGANESAN, 2017) que é o antecedente à evidenciação de dados nas bases de instituições públicas, bem como também em bases de dados de instituições privadas.

4. PROCESSO DE MINERAÇÃO DE DADOS

Conforme obra de Turban et al. (2008, 153-6) Data Mining (DM) é um termo usado para descrever a descoberta automática de informações em bancos de dados e seus algoritmos podem ser divididos em 4 categorias consolidadas:

- 1 - Classificação
- 2 - Agrupamento
- 3 - Associação
- 4 - Descoberta de sequência

Conforme Turban et al. (2009, pág. 155-6), Classificação ou Indução Supervisionada talvez seja a mais comum de todas as atividades de DM, pois o objetivo é analisar dados históricos e gerar automaticamente um modelo que possa prever comportamento futuro. Esse tipo de objetivo é de uma riqueza em gestão muito significativo pois pode-se aumentar a receita, reduzir custos, reduzir estoques (sistema *lean manufacturing* na produção) mas também mitigar riscos e antecipar aspectos de melhor prestação de contas. Se utiliza aqui algoritmos de Redes Neurais (cada nós da rede fazendo seus próprios cálculos, melhor denominado no algoritmo Multiplayer Perceptron), Árvores de Decisão (hierarquia de declarações se-então para dados categorizados e intervalares, melhor denominado com o Algoritmo C4.5 e Regras de Se-Então que necessariamente não precisam ter uma estrutura em árvore.

No Agrupamento o banco de dados é dividido em segmentos cujos membros compartilham qualidades semelhantes e as redes neurais podem ser utilizadas de maneira muito eficiente e cujo objetivo inclui uma otimização para criar grupos em que seus membros dentro de cada grupo tenham semelhança máxima, mas que os membros fora dos grupos tenham semelhança mínima.

Na Associação se estabelecem relações entre itens que ocorrem juntos em um determinado registro em que se pode aumentar significativamente as vendas, mas também melhorar em muito a prestação de contas quando se associa itens para comprovar fraudes financeiras. Na Descoberta de sequência se identifica Associações em longo prazo analisando-se frequência com eventos e assim compreendendo o comportamento de compras para exemplificar.

Mineração de dados permite identificar padrões, predizer situações (WAN et al., 2016) e realizar aprendizado de máquina entre outros recursos de igual robustez combinando técnicas estatísticas, inteligência artificial e regras de negócios muito aplicadas no setor financeiro, de saúde e vendas (FARAZZMANESH e HOSSEINI, 2017). Assim, pode-se analisar rede B2B de valor de clientes, bem como em auditorias diversas, inclusive para avaliação de níveis de serviços gerados por instituto de pesquisa público coreano (CHOI et al., 2017) com o uso intensivo de *random forest*

regression e mesmo aumentando garantias de privacidade de conduta em negócios online (PAYNE, LANDRY e DEAN, 2015).

Sendo assim esse tipo de modelagem permite gerar maior prestação de contas na empresa, maior prestação de contas com menores riscos em aplicações empresariais diversas e até mesmo no *e-commerce* tão difundido atualmente com aquisição de passagens aéreas automatizadas e reserva de hotéis por todo o mundo.

Esses enfoques necessitam integrar muito bem dados como analisado em Giraldo, Jiménez e Tabares (2017) as relações de DM x Business Intelligence (WANG e WANG, 2008) x Data Warehouse, alcançando inclusive uma aprendizagem de máquina e geração de conhecimento (HEINRICHS e LIM, 2003) que conectam social networks (TRANDAFILI e BIBA, 2013).

O método DM se utiliza de algoritmos para extração de modelos dos dados, conhecimento e aprendizado de máquina e está fundamentado em Sistema de Apoio à Decisão (SAD) ou *Decision Support System* (DSS), sistema esse de apoio à resolução de problemas pouco ou não estruturados (KEEN e SCOTT-MORTON, 1978; SPRAGUE e CARLSON 1982; SPRAGUE e WATSON, 1989).

Assim, DM é um método de extração de modelos de dados, geração de conhecimento e aprendizado automático de máquina que permite prover o decisor de informações, conhecimento e também análise de diferentes alternativas considerando associações, agrupamentos e tendências (WITTEN e FRANK, 2005) devido a padrões não usuais destas bases de dados, encontrados em Data Warehouse (DW) defendidas em Kimbal (1998) e Inmon (1997) ou também caracterizado como Big Data, normalmente não previstas em sistemas de informações gerenciais usuais.

O enfoque *Knowledge Discovery in Database* (KDD) se aplica para análise inteligente através de avaliação automática ou semi-automática desses grandes repositórios de dados reforçando o objetivo de encontrar informações relevantes relacionadas a erros ou não conformidades, inclusive fraudes (WITTEN e FRANK, 2005). Também se avalia a qualidade de modelos construídos podendo ser avaliados por diferentes medidas de qualidade de conhecimento descoberto (HUANG e LING, 2005) em que se reduza não conformidades com melhores padrões de eficiência (MONEDERO et al., 2012).

As etapas podem ser realizadas de acordo com Fayyad (1996) como segue:

- Seleção dos dados
- Pré-processamento e limpeza dos dados
- Transformação dos dados
- Data Mining
- Interpretação / Avaliação

Estas etapas estão dentro de um ciclo de design com ênfase em gestão, ou seja, DM para Management. Segundo Sharma, Osei-Bryson e Kasper (2012), a descoberta de conhecimento através do método de Mineração de Dados é um processo de múltiplas fases que inclui:

- O entendimento sobre negócio
- Preparação dos dados
- Modelagem

- Avaliação
- Implantação

O processo KDD é iterativo e complexo, uma vez que cada fase envolve múltiplas tarefas, e existem numerosas dependências que integram essas fases sustentadas por bases de dados e os respectivos algoritmos de DM.

A descoberta de conhecimento em bases de dados não é um processo trivial de identificação de novas informações, válidas e potencialmente úteis com o objetivo de gerar de forma eficiente resultados que possam ser visualizados e interpretados através da interação homem-máquina (FAYYAD, PIATETSKY-SHAPIO e PADHRAIC, 1996). Porém, segundo Williams (2012), o método de DM é muito utilizado como uma ferramenta crítica para a coleta de informações relevantes para as instituições bem como na identificação de riscos operacionais. Também salienta-se aqui a aplicação e diálogo com auditorias diversas, inclusive para avaliação de níveis de (CHOI et al., 2017) e mesmo aumentando garantias de prestação de contas.

Desta forma, é importante que este trabalho analise a robustez desse tipo de abordagem de Data Science com o uso de DM através do *software* Weka, sendo que esse enfoque busca proporcionar uma maior aproximação do ambiente de gestão a esse enfoque de análise de dados, mas com maior protagonismo do gestor, ou seja, do decisor. Assim, isso proporciona uma estruturação de classificação aplicada ao *accountability* para avaliar dados com maior confiança e afirmação para melhoria do processo decisório (CHANNUNTAIPAT, 2018), bem como em reduzir possibilidades de fracassos ou impotências da mesma quando esta estiver integrada com contabilidade e governança (RADCLIFFE, SPENCE e STEIN, 2017).

5. ETAPAS DA PESQUISA

Para operacionalização deste trabalho se apresenta na continuação as etapas de Data Mining que estão relacionadas à técnica de coleta e análise de dados, envolvendo operacionalmente também as 5 etapas (FAYYAD, 1996) mais operacionais relacionadas a: 1) Seleção dos dados, 2) Pré-processamento e limpeza dos dados, 3) Transformação dos dados, 4) Data Mining e 5) Interpretação e avaliação dos resultados.

Para a técnica de Análise de Dados utiliza-se de algoritmos de mineração de dados direcionados para a modelagem em Aprendizagem Supervisionada considerando diferentes etapas (HERNÁNDEZ, RAMÍREZ e FERRI, 2004) como:

1 - Identificação das principais atividades evidenciadas na base de dados de processos tecnológicos. No caso aqui se utilizou de base de dados envolvendo 548 empresas, este desenvolvido desde o ano de 2012.

2 - Direcionamento para classificação seletiva de requisitos de Prestação de Contas em base de dados no Portal da Transparência com realização do processo de KDD com classificação, agrupamento, regras associativas, ou outra considerando aprendizados supervisionados.

Estas etapas podem seguir por caminhos como apresentado em modelo *Cross-Industry Standard Process for Data Mining* (CRISP-DM), representado na Figura 1.

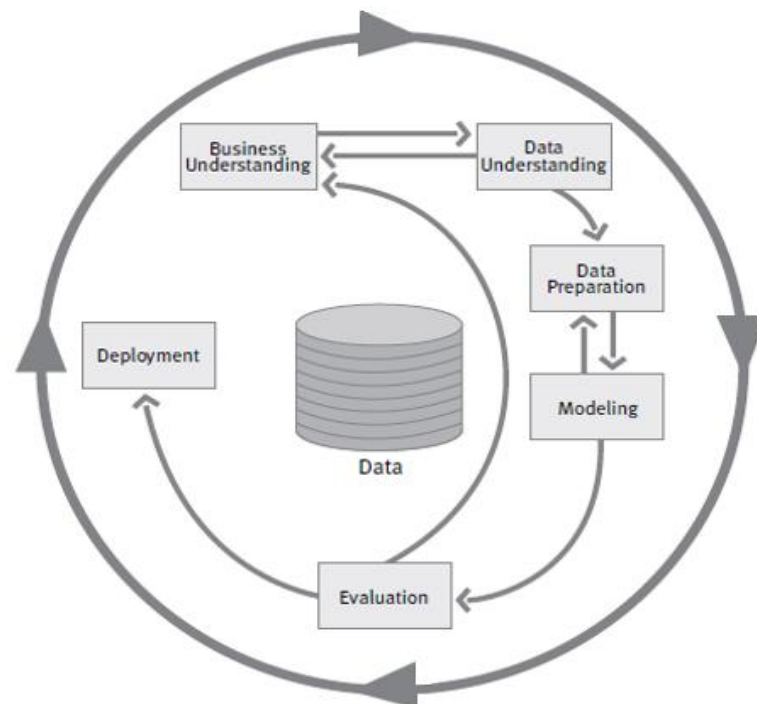


Figura 1 - Fases do modelo de referência CRISP-DM.

Conforme TURBAN, SHARD, ARONSON, KING (2009, p. 173) “*este modelo foi desenvolvido destacando-se as fases de entendimento da necessidade do negócio e preparação de dados como as mais exaustivas*”. Chapman et al. (2000) descrevem as seguintes etapas aplicadas à metodologia:

- Requisitos de negócio: Etapa inicial de entendimento dos objetivos e requisitos do projeto a partir da perspectiva de negócio em que obtendo-se esse entendimento pode-se definir problema e plano de atuação.

- Necessidade de dados: Com a coleta de dados é possível então qualificar e obter os primeiros resultados como é a identificação de clusters e mesmo diálogo com formação de hipóteses para resolução de algum problema em que a organização não perceba que esteja ocorrendo.

- Tratamento de dados: Transformação do dado bruto em conjunto final de dados através de modelagem que se classifique e se selecione atributos, realizando “limpeza” dos mesmos e sua transformação. A partir dessa etapa também pode-se realizar *Business Intelligence*.

- Modelagem: Seleção e aplicação de várias técnicas com ajuste dos parâmetros para otimizar os valores e validá-las em rodadas até que as mesmas sejam adequadas ao problema que se queira solucionar.

- Avaliação: Talvez seja a etapa mais crítica para o gestor e o distanciamento ao técnico de DM pois por vezes o DM pode não estar alinhado como se espera e encontrar obviedades na empresa. Isso pode ocorrer por seleção de técnicas e algoritmos não adequado àquela realidade.

- Implementação: Conversão do conhecimento de tal forma que o decisor possa se apropriar, utilizar, melhorar e no caso desse enfoque, proporcionar à instituição

seja pública ou privada, uma melhor prestação de contas. Considera-se aqui também uma interface *user friendly* e recursos para gerar consultas de maneira dinâmica.

6. RESULTADOS

Nesta investigação operacionalmente se considerou os domínios de *Planning and Organization* (PO), *Acquisition and Implementation* (AI), *Delivery and Support* (DS), mas se centrou esforços de geração de análise diversas como representado pelo *output* de Weka e com validações em trabalhos subsequentes desta mesma pesquisa.

Foi possível então inicialmente gerar tela de leitura do arquivo em que imediatamente o Weka disponibiliza algumas estatísticas descritivas conforme Figura 2 da continuação.

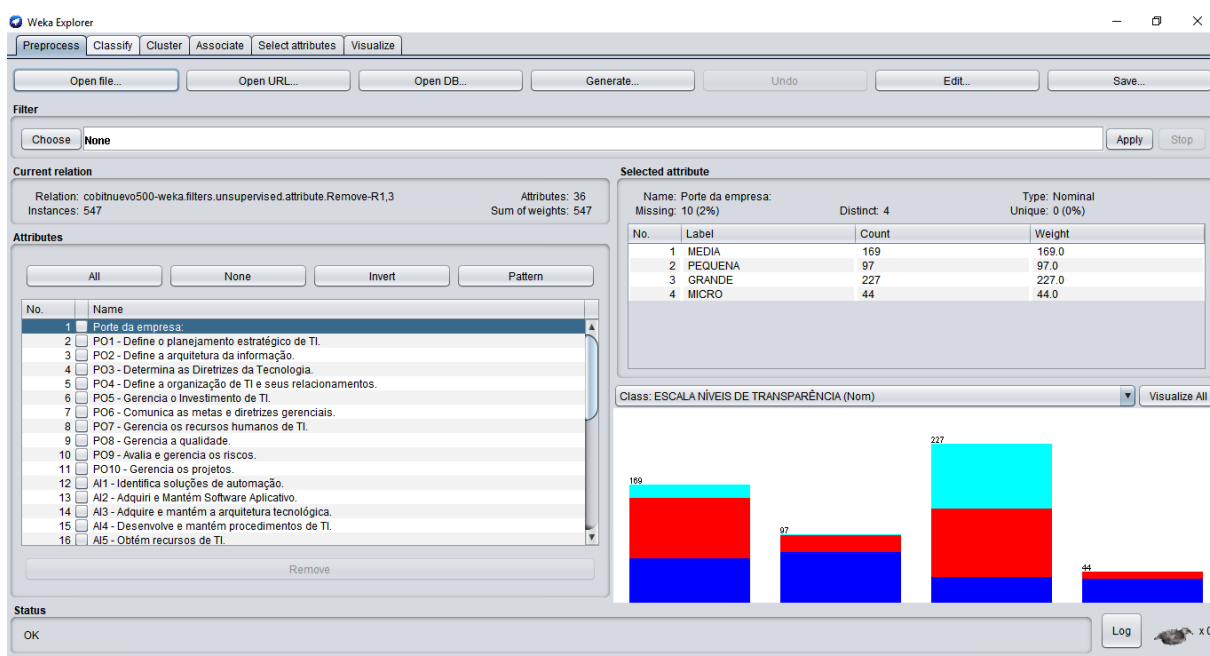


Figura 2 - Análise de processos tecnológicos com técnicas de Data Minig.

A Figura 2 apresenta os processos tecnológicos analisados e de imediato as técnicas de mineração de dados como Classificação, Clusterização ou Agrupamento, Associação e Seleção de atributos.

Pode-se também visualizar o que ocorre com a base de dados e estabelecer compreensão de dados outliers em que diversos estudos acabam desconsiderando-os. Em Mineração de dados o que destoa pode apresentar um significado alto em gestão como algum tipo de fraude. Não é o caso aqui pois não foi o foco deste trabalho, mas contribui para averiguar o porquê que certos dados estão se distanciando de por exemplo um agrupamento de dados.

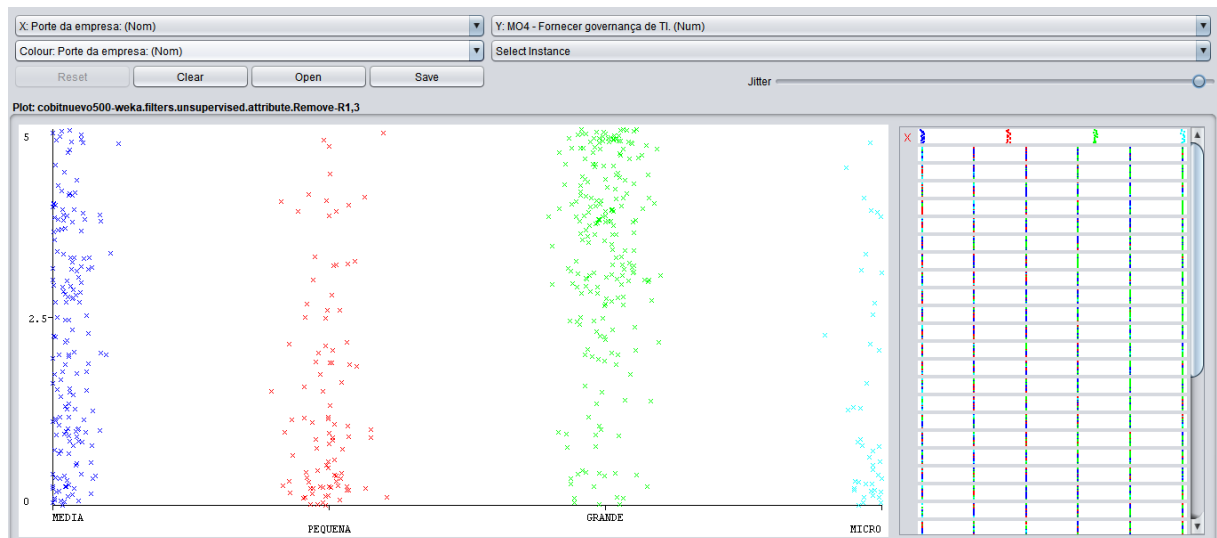


Figura 3 - Análise de cluster.

Esse tipo de funcionalidade do *software* pode indicar uma mudança de comportamento esperado com os padrões pré-estabelecidos de clusterização em uma base de dados através de uma visualização. É apresentado na Figura 4 uma forma deste tipo de Agrupamento ou clusterização que o Weka disponibiliza ao usuário, sendo que em sua primeira parte são listados os processos tecnológicos, logo os 6 clusters e finalmente uma consolidação dos mesmos.

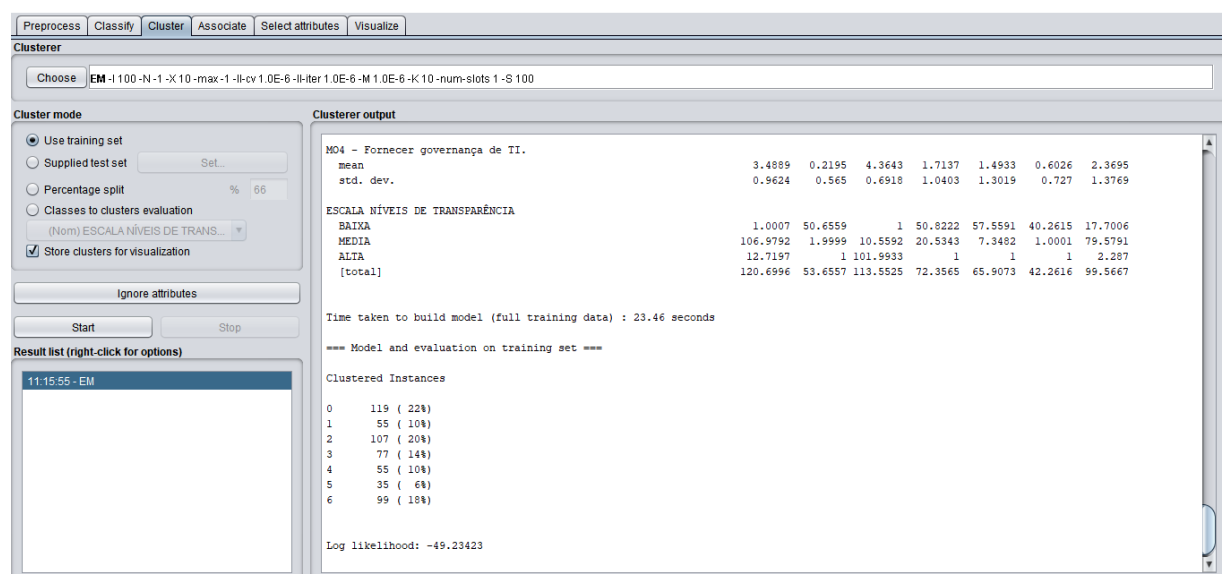


Figura 4 - Exemplo de Clusterização.

A figura 4, anteriormente disponibilizada é uma exemplificação em que posteriormente as análises seriam avaliadas e logo validadas ou não. Já na Figura 5 da continuação pode-se verificar um exemplo de árvore de decisão utilizando-se J48 que é o algoritmo para este tipo de técnica.

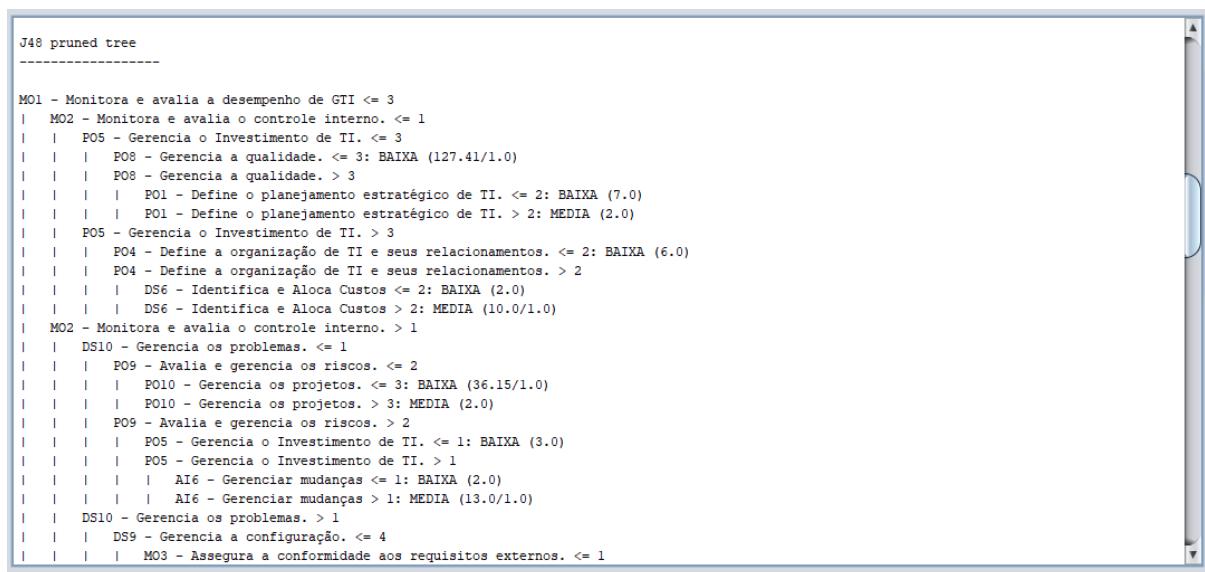


Figura 5 - Exemplo de parte de uma árvore de decisão.

Figura 5 mostra parte de uma árvore de decisão, utilizando-se J48 que é o algoritmo para este tipo de técnica de classificação. O Weka apresenta muitos recursos para realizar a mineração de dados, é didático e com a grande vantagem em ser *software* de código aberto, ou seja, acessível a todos os usuários. Gradativamente se pretende aproximar mais usuários finais ou gestores do uso destas poderosas técnicas e incorporar os algoritmos de descoberta de conhecimento ao seu dia a dia. O uso dos mesmos exige que de maneira rápida e validada se possa compreender melhor o que dizem os dados nas instituições e assim prestar melhores contas de suas atividades.

7. CONCLUSÕES

Este trabalho de Data Science com Data Mining objetivou aproximar didaticamente gestores e pesquisadores de gestão a esta importante técnica, porém sem perder robustez e ilustrando com uma promoção de governança de TI (MO4) considerando o *framework* Cobit da reconhecida ISACA. Sendo assim, através de níveis consolidados de maturidade do Cobit, estes estruturados em domínios e processos, respondidos junto a 548 empresas brasileiras, pôde-se simular algumas situações e posicionar para desenvolver uma melhor prestação de contas institucional, seja pública ou privada.

A análise didática realizada permite aproximar as técnicas de Data Mining ao contexto gerencial. Tem sido possível contribuir para que as empresas entendam como melhorar a prestação de contas, aperfeiçoando uma estrutura de governança tecnológica que reduza as assimetrias de informação e melhore a transparência e a qualidade das informações.

Os avanços nesta linha de pesquisa, através de técnicas de mineração de dados combinadas com árvores de decisão e aprendizado de máquina, nos permitem construir modelos computacionais com uma abordagem matemática para a descoberta de conhecimento para promover a Governança de TI nas empresas. Da mesma forma, essas abordagens contribuem para uma cultura de melhoria dos processos de negócios, considerando de forma mais sistemática os requisitos de auditoria alinhados à governança corporativa e aos processos de TI nas empresas.

REFERÊNCIAS

AGUIAR, J.; PEREIRA, R.; VASCONCELOS, J.; BIANCHI, I. (2018). An overlap less incident management maturity model for multi-framework assessment (ITIL, COBIT, CMMI-SVC). **Interdisciplinary Journal of Information, Knowledge, and management**, 13, 137-163. DOI: <https://doi.org/10.28945/4083>.

ALHINAI, Y.; AL-BADI, A.; AL-HARTHI, I.; AL-SALTI, Z. (2016). Rethinking IT-governance: Analytics review of IT Governance for social media based on the Cobit standar. **International Journal of Services, Economics and Management**. 7(2-4), 124-153.

ALHURAIBI, A. (2017). **From IT-business strategic alignment to performance: A moderated mediation model of social innovation, and enterprise governance of IT**. TICC Series. Dutch Research School for Information and Knowledge Systems. Tilburg University.

ALKHALDI, F. M.; HAMMAMI, S. M.; UDDIN, M. A. (2017). Understating value characteristics toward a robust IT governance application in private organizations using COBIT framework. **International Journal of Engineering Business Management**, 9, 1-8. DOI: 10.1177/1847979017703779.

ALOM, M. (2018) Proactive transparency and outward accountability of frontline public bureaucracies: An integrated model. **International Journal of Productivity and Performance Management**, 67(4), 611-628.

ATKINS, J.; MAROUN, W. (2018). Integrated extinction accounting and accountability: FAYYAD, U. M. (1996). Data mining and knowledge discovery: Making sense out of data. **IEEE Intelligent systems and Their Applications**, 11(5), 20-25.

FAYYAD, U. M.; SHAPIRO, G; PADHRAIC, S. (1996). From Data Mining to Knowledge Discovery in Databases. **AI Magazine**. 17(3), 37-54.

building an ark. **Accounting, Auditing and Accountability Journal**, 31(3), 750-786.

AWAD, A. I. (2018). Introduction to information security foundations and applications. In **Information Security: Foundations, Technologies and Applications**. The Institution of Engineering and Technology (IET), 3-11.

BRANDES, L.; DARAI, D. (2017) The value and motivating mechanism of transparency in organizations. **European Economic Review**, 98(C), 189-198. DOI: 10.1016/j.euroecorev.2017.06.014.

CAMPREGHER, G.; LONGONI, L. S. (2017). A natureza humana do comportamento individual nos primórdios do pensamento econômico: uma comparação entre Hume, Smith e Bentham. **Economia e Sociedade**, 26(1), 111-139, DOI: [10.1590/1982-3533.2017v26n1art4](https://doi.org/10.1590/1982-3533.2017v26n1art4).

CARVALHO, M.; ROMÃO, M.; FAROLEIRO, P. (2016). Governança e projetos de TI: a integração COBIT 5 e PMBOK. **16ª. Conferência da Associação Portuguesa de Sistemas de Informação**. Porto, Portugal, DOI: 10.18803/capsi.v16.84-104.

CHANNUNTAIPAT, C. (2018). Assurance for service organisations: Contextualising accountability and trust. **Managerial Auditing Journal**, 33(4), 340-359.

CHAPMAN, P.; CLINTON, J.; KERBER, R.; KHABAZA, T.; REINARTZ, T.; SHEARER, C.; WIRTH, R. (2000). **CRISPDM 1.0 step-by-step data mining guide**. Technical report, CRISP-DM.

CHOI, J.; KIM, B.; HAHN, H.; PARK, H.; JEONG, Y.; YOO, J.; JEONG, M. K. (2017). Data mining-based variable assessment methodology for evaluating the contribution of knowledge services of a public research institute to business performance of firms. **Expert Systems with Applications**, 84(C), 37-48. DOI: 10.1016/j.eswa.2017.04.057.

CUGANESAN, S. (2017). The design of performance budgeting processes and managerial accountability relationships. **Public Management Review**, 19(7), 954-971.

DE HAES, S.; VAN GREMBERGEN W. (2015). **Enterprise Governance of IT: Achieving Alignment and Value**. Springer.

DEBRECENY, R.; GRAY, G. (2013). IT Governance and Process Maturity: A Multinational Field Study **Journal of Information Systems**. 27(1), 157-188, DOI: 10.2308/isys-50418.

DHIMAN, A.; SEN, A.; BHARDWAJ, P. (2018). Effect of Self-Accountability on Self-Regulatory Behaviour: A Quasi-Experiment. **Journal of Business Ethics**, 148(1), 79-97.

DU RIETZ, S. (2018). Information vs knowledge: Corporate accountability in environmental, social, and governance issues. **Accounting, Auditing and Accountability Journal**, 31(2), 586-607.

FARAZZMANESH, F.; HOSSEINI, M. (2017). Analysis of business customers' value network using data mining techniques. **Journal of Information Systems and Telecommunication**, 5(3), 162-171.

FAYYAD, U. M. (1996). Data mining and knowledge discovery: Making sense out of data. **IEEE Intelligent systems and Their Applications**, 11(5), 20-25.

FAYYAD, U. M.; SHAPIRO, G.; PADHRAIC, S. (1996). From Data Mining to Knowledge Discovery in Databases. **AI Magazine**. 17(3), 37-54.

FERRY, L.; MURPHY, P. (2018). What about Financial Sustainability of local Government! A critical review of accountability, transparency, and public assurance arrangements in England during Austerity. **International Journal of Public Administration**, 41(8), 619-629.

FILGUEIRAS, F. (2018). **Burocracias do controle, controle da burocracia e accountability no Brasil**. In Burocracia e Políticas Públicas no Brasil. RJ, Brasil: IPEA.

GANTMAN, S.; FEDOROWICZ, J. (2016). Communication and control in outsourced IS development projects: Mapping to COBIT domains. **International Journal of Accounting Information Systems**, 21(C), 63-83. DOI: 10.1016/j.accinf.2016.05.001.

GAO, F.; RAU, P. L. P.; ZHANG, Y. (2018). Perceived Mobile Information Security and Adoption of Mobile Payment Services in China. In **Mobile Commerce: Concepts, Methodologies, Tools, and Applications**. USA: IGI Global.

GIRALDO, J. C.; JIMÉNEZ, J.; TABARES, M. S. (2017). Model to optimize business processes management with the integrating mining of processes and business intelligence in data warehouse. **Espacios**, 38(2), art. no. 9.

GIRIJA, N.; SRIVATSA, S. K. (2006). A research study: Using data mining in knowledge base business strategies. **Information Technology Journal**, 5(3), 590-600.

HASBINI, M. A.; ELDABI, T.; ALDALLAL, A. (2018). Investigating the information security management role in smart city organisations. **World Journal of Entrepreneurship, Management and Sustainable Development**, 14(1), 86-98.

HEINRICHS, J.H.; LIM, J.-S. (2003) Integrating web-based data mining tools with business models for knowledge management. **Decision Support Systems**, 35(1), 103-112.

HERNÁNDEZ, J.; RAMÍREZ, M. J.; FERRI, C.; (2004). **Introducción a la Minería de Datos**. España: Prentice-Hall.

HINA, S.; DOMINIC, P. D. D. (2018). Information security policies' compliance: a perspective for higher education institutions. **Journal of Computer Information Systems**, 1-11, DOI: 10.1080/08874417.2018.1432996.

HOLLAND, D.; KRAUSE, A.; PROVENCHER, J.; SELTZER, T. (2018). Transparency tested: The influence of message features on public perceptions of organizational transparency. **Public Relations Review**, 44(2), 256-264. DOI: 10.1016/j.pubrev.2017.12.002.

HUANG, J.; LING, C. X. (2005). "Using AUC and accuracy in evaluating learning algorithms." **IEEE Transactions on Knowledge and Data Engineering**. 17(3): 299-310.

INMON, W. (1997). **Como construir o Data Warehouse**. Rio de Janeiro: Campus.

INSTITUTO BRASILEIRO DE GOVERNANÇA CORPORATIVA (IBGC) (2009). **Código das Melhores Práticas de Governança Corporativa**. 4. ed. São Paulo. Disponível em: <http://www.ibgc.org.br/CodigoMelhoresPraticas.aspx> . Acesso em: 15 jan, 2013.

ISACA (2012). **COBIT 5 – Modelo corporativo para governança e gestão de TI das organizações**. E-BOOK, Rolling Meadows, IL, EUA. Disponível em: http://www.cefet-rj.br/attachments/article/2870/Cobit_5_pt-br.pdf. Acesso em 12 jun. 2018.

ISHAQ, A.; MUKHTAR, M.; WAHYUDI, M.; INDRIANI, K. (2017). Information technology governance using Cobit 4.0 domain delivery support and monitoring evaluation. **Journal of Theoretical and Applied Information Technology**, 95(20), 5304-5315.

ITGI (2003). **Board briefing on IT Governance**. IL, USA: IT Governance Institute.

JAIN, S.; JAIN, S. (2018). Power distance belief and preference for transparency. **Journal of Business Research**, 89, 135-142. DOI: 10.1016/j.jbusres.2018.04.016.

KEEN, P.; SCOTT-MORTON, M. (1978). **Decision Support System**. Reading, MA: Addison-Wesley Publishing Company.

KIMBALL, R. (1998). **The Data Warehouse Toolkit - Técnicas para a Construção de Data Warehouses Dimensionais**. São Paulo: Makron Books.

KNUPLESCH, D.; REICHERT, M. (2017). A visual language for modeling multiple perspectives of business process compliance rules. **Software & Systems Modeling**, 16(3), 715-736.

KRAFT, B.; WOLF, S. (2018). Through the Lens of accountability: Analyzing Legitimacy in environmental governance. **Organization and Environment**, 31(1), 70-92.

LEE, H.; LEE, H.-L.; WANG, C. (2017). Engagement partner specialization and corporate disclosure transparency. **International Journal of Accounting**, 52(4), 354-369.

LÓPEZ DE MANTARAS, R.; ARMENGOL, E. (1998). Machine learning from examples: Inductive and Lazy methods. **Data & Knowledge Engineering**, 25(1-2), 99-123. DOI: 10.1016/S0169-023X(97)00053-0.

MCCUE, C. (2007). **Data Mining and Predictive Analysis**. Elsevier.

MITCHELL, T. (1997). **Machine Learning**. NY, USA: McGraw-Hill.

MONEDERO, I.; BISCARRI, F.; LEÓN, C.; GUERRERO, J.I.; GONZÁLEZ, R. (2012). Decision system based on neural networks to optimize the energy efficiency of a petrochemical plant. **Expert Systems with Applications**, 39(10), 9860-9867.

NAJMI, K.; KADIR, A.; KADIR, M. (2018). Mediation effect of dynamic capability in the relationship between knowledge management and strategic leadership on organizational performance accountability. **International Journal of Law and Management**, 60(2), 517-529.

NEI, K. S.; FOSTER, J. L.; NESS, A. M.; NEI, D. S. (2018). Rule breakers and attention seekers: Personality predictors of integrity and accountability in leaders. **International Journal of Selection and Assessment**, 26(1), 17-26. DOI: 10.1111/ijasa.12201

NICOL, O. (2018). No body to kick, no soul to DAM: Responsibility and accountability for the financial crisis (2007-2010). **Journal of Business Ethics**, 151(1), 101-114.

ODONGO, N.; WANG, D. (2018). Corporate responsibility, ethics and accountability. **Social Responsibility Journal**, 14(1), 111-122.

OECD (2015). **G20/OECD principles of corporate governance**. Paris, France: OECD Publishing. DOI: 10.1787/9789264236882-en

OXELHEIM, L. (2018). Optimal vs satisfactory transparency: The impact of global macroeconomic fluctuations on corporate competitiveness. **International Business Review**, 28(1), 190-206. DOI: 10.1016/j.ibusrev.2018.05.011.

PAPAZOV, E.; MIHAYLOVA, L. (2015). Organization of Management Accounting Information in the Context of Corporate Strategy. **Procedia - Social and Behavioral Sciences**, 213, 309-313. DOI: 10.1016/j.sbspro.2015.11.543.

PARK, M.; CHAI, S. (2018). Internalization of Information Security Policy and Information Security Practice: A Comparison with Compliance. **Proceedings of the 51st Hawaii International Conference on System Sciences**.

PAYNE, D.; LANDRY, B.J.L.; DEAN, M.D. (2015). Data mining and privacy: An initial attempt at a comprehensive code of conduct for online business. **Communications of the Association for Information Systems**, 37(1), art. no. 34, 717-732.

PROVOST, F.; FAWCETT, T. Data Science para Negócios: **O que você precisa saber sobre mineração de dados e pensamento analítico de dados**. Alta Books.

PUTRI, M.A.; LESTARI, V.A.; AKNURANDA, I. (2017). Audit of information technology governance using COBIT 4.1: Case study in PT. XY. (2017) **Internetworking Indonesia Journal**, 9(1), 47-52.

RADCLIFFE, V.; SPENCE, C.; STEIN, M. (2017). The impotence of accountability: The relationship between greater transparency and corporate reform. **Contemporary Accounting Research**, 34(1), 622-657.

RIOS, O. K. L.; DE ALMEIDA TEIXEIRA FILHO, J. G.; DA SILVA RIOS, V. P. (2017). Melhores práticas do COBIT, ITIL e ISO/IEC 27002 para implantação de política de segurança da informação em Instituições Federais do Ensino Superior. **Revista Gestão & Tecnologia**, 17(1), 130-153.

ROMZEK, B.; INGRAHAM, P. W. (2000). Cross pressures of accountability: initiative, command, and failure in the ron brown plane crash. **Public Administration Review**, 60(3), 240-253. DOI: 10.1111/0033-3352.00084.

RYU, S.; CHANG, Y. (2017). Accountability, Political Views, and Bureaucratic Behavior: A Theoretical Approach. **Public Organization Review**, 17(4), 481-494.

SEIBERT, R. M.; MACAGNAN, C. B. (2015). Evidenciação das Instituições Comunitárias de Ensino Superior: Um estudo sob a perspectiva dos públicos de interesse. **Contextus - Revista Contemporânea de Economia e Gestão**, 13(2), 176-209.

SHAMBAUGH, G. E.; SHEN, E. B. (2018). A clear advantage: The benefits of transparency to crisis recovery. **European Journal of Political Economy**, 55(C), 391-416. DOI: 10.1016/j.ejpoleco.2018.03.002.

SHARMA, S.; OSEI-BRYSON, K. M.; KASPER, G. M. (2012). Evaluation of an integrated Knowledge Discovery and Data Mining process model. **Expert Systems with Applications**, 39(13), 11335-11348.

SHIRI, M. M.; AMINI, M. T.; RAFTAR, M. B. (2012). Data Mining Techniques and Predicting Corporate. **Interdisciplinary Journal of Contemporary Research in Business**, 3(2), 61-69.

SHMUELI, G.; KOPPIUS, O. R. (2011). Predictive analytics in information systems research. **MIS Quarterly**, 35(3) 553-572.

SINGH, V.; MARGAM, M. (2018). Information Security Measures of Libraries of Central Universities of Delhi: A Study. **DESIDOC Journal of Library & Information Technology**, 38(2), 102.

SPRAGUE, R.; CARLSON, E. (1982). **Building Effective Decision Support Systems**. NY, USA: Prentice-Hall.

SPRAGUE, R.; WATSON, H. (1989). **Decision support systems: putting theory into practice**. USA: Prentice-Hall.

STOUTHUYSEN, K.; SLABBINCK, H.; ROODHOOFT, F. (2017). Formal controls and alliance performance: The effects of alliance motivation and informal controls. **Management Accounting Research**, 37, 49-63. DOI: 10.1016/j.mar.2017.03.002.

SURYANI, M.; SASMITA, G.; PURNAWAN, I. Audit of accounting information system using COBIT 4.1 focus on deliver and support domain. 2015). **Journal of Theoretical and Applied Information Technology**, 78(3), 456-463.

TAN, P.; STEINBACH, M.; KUMAR, V. (2005). **Introduction to Data Mining**. Pearson Addison-Wesley.

TARMUJI, A.; SETIADI, T.; HANDAYANINGSIH, S.; LESTARI, J. (2017). Development of a customer relationship management model based on maturity level of Cobit 4.1: Case study of the cooperative section at department of industry, trade, cooperative, and small-medium enterprises, Yogyakarta province. **Asia-Pacific Journal of Science and Technology**, 22(2), art. no. APST-22-02-04, 6 p.

THOMANN, E.; HUPE, P.; SAGER, F. (2018). Serving many masters: Public accountability in private policy implementation. **Governance**, 31(2), 299-319.

TRANDAFILI, E.; BIBA, M. (2013). A review of machine learning and data mining approaches for business applications in social networks. **International Journal of e-Business Research**, 9(1), 36-53.

TURBAN, E.; SHARD, R.; ARONSON, J. E.; KING, D. (2009). **Business intelligence: um enfoque gerencial para a inteligência do negócio**. Porto Alegre: Bookman.

VALLE, M.; VARAS, S.; RUZ, G. (2012). Job performance prediction in a call center using a naive Bayes classifier. **Expert Systems with Applications**, 39(11), 9939-9945.

WAN, J.; YUE, Z.-L.; YANG, D.-H.; YU, Z.; JIAO, L.; ZHI, L.; LIU, J. (2016). Predicting non performing loan of business bank with data mining techniques. **International Journal of Database Theory and Application**, 9(12), 23-34.

WANG, H.; WANG, S. (2008). A knowledge management approach to data mining process for business intelligence. **Industrial Management and Data Systems**, 108(5), 622-634.

WEILL, P.; ROSS, J. (2005). **IT Governance: How Top Performers Manage IT Decision Rights**. USA: HBSP.

WILLIAMS J. (2012). Global trends and tribulations in mining regulation. **Global of Energy & Natural Resources law**, 30(4) 391-422.

WITTEN, I.; FRANK, E. (2005). **Data Mining: Practical Machine Learning tools and Techniques**. 2nd Edition. San Francisco, CA: Elsevier.

WU, X.; V. KUMAR, et al. (2008). Top 10 algorithms in data mining. **Knowledge and Information Systems**, 14(1), 1-37.